Journal of Biomedical Informatics 54 (2015) 167–173

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Prediction of clinical risks by analysis of preclinical and clinical adverse events

Matthew Clark

Elsevier Life Science Solutions, 1600 John F. Kennedy Blvd, Suite 1800, Philadelphia, PA 19103, United States

ARTICLE INFO

Article history:
Received 6 November 2014
Accepted 24 February 2015
Available online 5 March 2015

Keywords:
Bayesian
Translational medicine
Animal–human concordance
Adverse events
Drug risk
Likelihood ratio

ABSTRACT

This study examines the ability of nonclinical adverse event observations to predict human clinical adverse events observed in drug development programs. In addition it examines the relationship between nonclinical and clinical adverse event observations to drug withdrawal and proposes a model to predict drug withdrawal based on these observations. These analyses provide risk assessments useful for both planning patient safety programs, as well as a statistical framework for assessing the future success of drug programs based on nonclinical and clinical observations.

Bayesian analyses were undertaken to investigate the connection between nonclinical adverse event observations and observations of that same event in clinical trial for a large set of approved drugs. We employed the same statistical methods used to evaluate the efficacy of diagnostic tests to evaluate the ability of nonclinical studies to predict adverse events in clinical studies, and adverse events in both to predict drug withdrawal. We find that some nonclinical observations suggest higher risk for observing the same adverse event in clinical studies, particularly arrhythmias, QT prolongation, and abnormal hepatic function. However the lack of these events in nonclinical studies is found to not be a good predictor of safety in humans. Some nonclinical and clinical observations appear to be associated with high risk of drug withdrawal from market, especially arrhythmia and hepatic necrosis. We use the method to estimate the overall risk of drug withdrawal from market using the product of the risks from each nonclinical and clinical observation to create a risk profile.

1. Introduction

Nonclinical animal models have long been accepted as a means to examine both the efficacy and toxicity of drugs before administration to humans. The motivation has always been to reduce risk to humans by observing the outcomes in animals [1]. The desire for animal testing for toxicity was increased after public tragedies of the "Lash-Lure" case in which aniline eyelash dyes caused blindness, the toxicity associated with the formulation of sulfanilamide with ethylene glycol, and the teratogenic effects of thalidomide [2–4].

Animal models for toxicity have been shown to correctly represent human toxicities in many cases [5]. However there are relatively few statistical studies evaluating the concordance of nonclinical and clinical observations [6]. The study of 30 compounds in various species by Goldsmith et al found that animals well predicted the maximum tolerated dosages for clinical trials [7]. Fletcher examined 45 compounds and found a low concordance between animal and human adverse events [8]. The current canonical work in animal–human concordance for toxicity is the Olson study which examined animal and human toxicity of 150 compounds from a variety of therapeutic areas. In that work, the overall true positive animal–human concordance rate was 7% for rodent only, 36% for combination of rodent and non-rodent, 27% for single non-rodent species, and 70% for observation in any species [9]. Several other studies have been published since then with generally similar results [10–13]. However, the majority of the evidence for the efficacy of animal studies is based on true positives, with limited analysis of the false positives and false negatives [14,15].

Evaluation of the ability of animal models to predict human responses and toxicities is critical now that there are increasing pressures to reduce animal testing in favor of in-vitro and computational predictive methods [16]. In this work we compared data for drugs that have matched nonclinical and clinical data presented in FDA and EMEA submissions and analyze the results to measure the concordance between nonclinical and clinical adverse event observations. The human–animal concordance is measured using

E-mail address: m.clark@elsevier.com

http://dx.doi.org/10.1016/j.jbi.2015.02.008
1532-0464/

Bayesian statistical methods similar to those used to evaluate the efficacy of diagnostic tests. In addition, we study the relationship between the events and eventual drug withdrawal to look for observations that are statistically correlated to drug failure.

2. Methods

2.1. Data sources

The PharmaPendium database from Elsevier was used to summarize the adverse events each selected MedDRA adverse classes reported for each drug, as shown in Fig. 1 [17,18]. PharmaPendium contains data for 3815 drugs or drug formulations. The data is mined from FDA and EMEA documents released in connection with drug approval, and is supplemented by data from Mosby's Drug Consult and Meyler's Side Effects of Drugs [19,20]. Of the drugs that appear in the database, 102 have been withdrawn from the market or relabeled, providing a limited set of failure results for statistical analysis. In this case each approved formulation was considered separately since it is possible for drug combinations to have different adverse events than the drugs alone. Fig. 1 shows an example of the raw data from PharmaPendium that summarizes the number of times arrhythmias were reported for each drug. Each value is linked to full reports of each observation, linked to the original submission documents. Post-marketing adverse event reports were not used in this study; only those reported in controlled clinical studies were used.

In order to eliminate dependence on the number of submissions and clinical studies of each drug for various indications, each value in the table was translated to an indicator variable of 1, if there was an observation of an event in the category, and 0 if there were none. This was done to avoid using the raw counts which may be dependent on the number of studies performed for a particular drug. The indicator denotes that the drug is reported at least once to cause the effect. These indicator values were then used for the analyses.

	clinical observation	no clinical observation
nonclinical observation	a	c
no nonclinical observation	b	d

cells contain count of drugs in each category for the given class of adverse events

Fig. 2. 2 × 2 Contingency table used for statistical analysis.

2.2. Bayesian statistics

Bayesian statistics were used with a 2 by 2 contingency table to measure the relationship between two sets of observations – the relationships between nonclinical and clinical observations for adverse events or adverse event categories, and separately, combined nonclinical/clinical observations with drug withdrawal (see Fig. 2). We treat the nonclinical observation as a diagnostic test for the clinical observation and use the statistical methods developed for evaluation of the efficacy of diagnostic tests. The same analysis is applied to measure the relationship between adverse event observation and drug withdrawal.

The values in the 2 by 2 contingency table, which are counts of number of compounds in each of the four categories for a given biomedical observation or MedDRA class of observations, were generated as follows:

(a) Count of drugs for which the event was observed in both nonclinical and clinical studies – true positives.

Cardiac arrhythmias

Drugs *:

- View by drug class
- **Viewing by name**

	Preclinical Data view all 881	Clinical Data view all 26073	Post-Marketing Reports (AERS) view all 284208
1-13C-Caprylic Acid	2	no data	no data
Abacavir Sulfate	no data	3	244
Abacavir Sulfate; Lamivudine	no data	1	58
Abacavir Sulfate; Lamivudine; Zidovudine	no data	no data	71
Abarelix	2	no data	13
Abatacept	no data	11	129
Abciximab	no data	5	458
Abetimus Sodium	no data	1	no data
Abiraterone Acetate	2	70	76
Acamprosate Calcium	11	1	20
Acarbose	no data	1	43
Acebutolol Hydrochloride	1	41	60
Acecainide	1	1	no data

Fig. 1. PharmaPendium summary of adverse events.

(b) Count of drugs for which the event was observed in clinical, but not nonclinical – false negatives.
(c) Count of drugs for the event was observed in nonclinical studies but not clinical studies – false positives.
(d) Count of drugs without the adverse event observation in either nonclinical or clinical studies – true negatives.

In this study we use likelihood ratios to express the statistical connection between nonclinical and clinical observations [21]. The likelihood ratio has the advantage that it is independent from the prevalence (Bayesian prior probability) for each observation so that it is more comparable across different adverse events than the conditional probability or positive predictive value [22]. This is important because the overall prevalence of adverse events is relatively low, which results in a low sensitivity metric, as well as low positive predictive value. The likelihood ratio computed here represents the *change* in clinical risk when the adverse event is observed in a nonclinical study.

The sensitivity, $a/(a+b)$, is the ability of the nonclinical adverse even observation to predict that the adverse event will be observed clinically. The specificity, $d/(c+d)$, is the ability of the nonclinical result to predict that the adverse event *will not be observed* in a clinical study.

The positive likelihood ratio (LR+) is computed from the sensitivity and the specificity. The positive likelihood is sensitivity/(1 – specificity), and the negative likelihood is (1 – sensitivity)/specificity. Positive likelihood ratios greater than 1.0 indicate a higher likelihood that the clinical observation will be made if the nonclinical observation is made. The negative likelihood (LR−) refers to the likelihood that there will be no clinical observation if no nonclinical observation is made. LR− values close to 0.0 suggest that the lack of nonclinical observation will predict lack of clinical observation. A value close to 1.0 for either LR+ or LR− suggests that there is no change in risk for either an observation or lack of an observation. A subjective explanation of the significance of the values is given in Table 1 [22]. The *p*-values for the relationships in the 2×2 tables were computed using Fisher's exact test implemented in R [23]. The Fisher's test is considered better than the chi-square test of significance for the unbalanced data tables in this study. Only relationships with *p*-values less than 0.05 were retained for this work, with the majority being orders of magnitude less than this value.

Fig. 3 shows example data for the adverse event "arrhythmia". There were 55 drugs for which arrhythmia was reported in nonclinical and clinical studies, 14 for which there was arrhythmia in a nonclinical study but not in a clinical study, and 946 drugs where there was a clinical observation but no nonclinical observation. Finally, there are 2800 drugs with no arrhythmia reported in either study type. The total sums to the 3815 approved drugs and formulations considered for this study. With the example data in Fig. 3, the positive likelihood, LR+, is 11.0, suggesting that if arrhythmia is observed in a nonclinical study there is high likelihood of observing it in human, 11 times that of observing it without a nonclinical observation. The negative likelihood, LR-, is 0.95, suggesting that the lack of observation of arrhythmia in nonclinical studies only slightly lowers the likelihood of seeing it in clinical studies. Analyses were carried out to examine the observations grouped in various MedDRA classifications of events, as well as for specific observations such as QT prolongation.

Table 1
Subjective interpretation of likelihood ratios (Ref. [22]).

LR+	LR−	Interpretation
>10	<0.1	Large and often conclusive shifts in probability
5–10	0.1–0.2	Moderate shifts in probability
2–5	0.2–0.5	Small, but sometimes important, shifts in probability
1–2	0.5–1	Alters probability to a small, and rarely important, degree

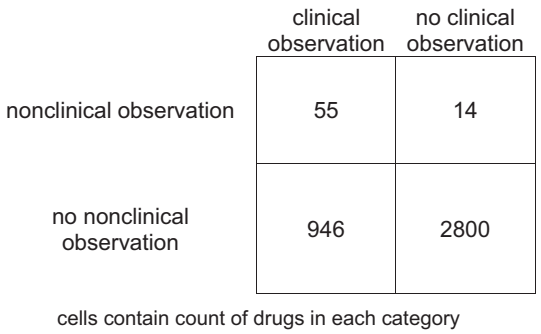

Fig. 3. The 2×2 contingency table used to compute likelihood ratios for observing an arrhythmia in clinical study if observed in a nonclinical study.

For prediction of drug withdrawal the same analyses were used, however the categories used in the 2×2 table were "withdrawn" and "not withdrawn" on the horizontal access and "adverse event observed" and "adverse event not observed" on the vertical access. This was used to compute the likelihood ratio of withdrawal for each drug/adverse event pair. As with the previous analysis only relationships with greater than 95% confidence as computed by the Fisher's exact test were retained.

2.3. Assumptions

Several important assumptions must be made about the uniformity of the data. We assume that the dosages in regulatory submissions for nonclinical adverse events are relevant to human doses. This work does not compare dosing or plasma concentration between nonclinical and clinical studies.

The reports of nonclinical toxicity are assumed to be only the 'significant' reports. That is they have been curated by the medical team to remove events caused by very high doses in safety studies. The assumption is that since these were successful clinical trials the animal and human doses reported are both relevant for the studies undertaken.

We also assume that the patient safety plan is monitoring for expected events; however, some unexpected events may not be monitored. These idiosyncratic events may be the cause of withdrawal or drug failure and could skew the statistics.

3. Results and discussion

3.1. Limitations and assumptions

3.1.1. Limitations

There are many issues complicating a statistical study of this nature. While it provides some guidance for the risks for clinical trials associated with nonclinical observations, some limitations must be noted.

The drugs represented in regulatory approval data are not a random sample. They represent the culmination of a long elimination process. Drugs with serious nonclinical events have been eliminated. Drugs with serious clinical events have also been eliminated due to discontinued trials and withdrawal of applications and that data is not available from regulatory documents. This process has removed many "true positive" signals from the statistics. However, at the same time it may have removed "false positive" drugs abandoned due to caution from animal effects that would not be observed in humans. In general the statistical conclusions here may not apply to preclinical compounds. These risk factors

may best apply to drugs that have advanced to at least phase II clinical trials.

The lack of data on the degree of the reported events is also a limitation of this analysis. A given clinical event could be of a minor or devastating degree. This limitation is ameliorated somewhat by the hope that the set of approved drugs has removed profound adverse events, and that only medically meaningful events are reported in approval documents.

The major reasons for drug withdrawal, or relabeling, are liver and cardiovascular issues, as shown in Table 2. Therefore development teams focus on these areas as possible roadblocks in the development process. This focus may increase the statistical reliability of these nonclinical–clinical relationships over the other less common issues. This work did not systematically measure the statistical significance of all possible observations. A more inclusive study may result in discovery of stronger markers for clinical trial risk factors.

3.2. Predictive relationship of nonclinical to clinical adverse events

The statistical analysis comparing adverse event observations in nonclinical and clinical studies is presented in Tables 3–5. Statistical analysis relating either nonclinical or clinical observations to drug withdrawal are shown in Tables 6 and 7. In this study the statistics all but one category includes over three thousand compounds, as compared to the 150 compounds considered in the Olson study. Negative likelihood, the ability to predict that if an observation is not made in a nonclinical study, it will not be observed in clinical studies provides the probability of "safety" for that observation. Positive likelihood is a prediction that if the observation is made in a nonclinical study, it will appear in a human clinical study. The relationships are not generally symmetric; one often observes that a high LR+ is not paired with a low LR−; meaning that the presence of a nonclinical observation may be strongly related to high risk in humans, but the lack of the same nonclinical observation does not generally imply safety in humans.

3.2.1. Cardiac issues

The results for cardiac issues are presented in Table 3. Long QT is a biomarker for a rare but potentially fatal ventricular arrhythmia – Torsades de pointes [25]. Therefore, there is a substantial effort in the early drug development process to detect drug-induced QT prolongation using appropriate animal models and accurately predict the risk of long QT Interval in humans.

In this analysis, the likelihood ratio for observing QT elongation in human, if observed in nonclinical models is 10.70, which suggest that observation of QT elongation in nonclinical studies predicts substantial human risk. The negative concordance, also of high interest, is 0.86. This value suggests that the lack of observation of QT prolongation in nonclinical studies does not inform one of the risks to human (based on statistics from this data). This is because there are 319 drugs where QT prolongation is observed in clinical studies without a nonclinical observation. Summing the count of drugs with any adverse event in the more inclusive MedDRA category of arrhythmia provides a similar statistic with a significant likelihood ratio of 5.84.

In general, the lack of a cardiac nonclinical observation does not appear to reduce the risk of it appearing in a clinical study. For example even though nonclinical observation of arrhythmia increases risk by a factor of 11, the lack of seeing arrhythmia unfortunately has no statistical implication for safety in humans in this data set because the LR− value is close to 1.0, at 0.95.

Table 3 also illustrates the difference between LR and sensitivity/specificity. For each item the sensitivity value is quite low,

Table 2
The top 14 systems affected by adverse reactions to 284 drugs withdrawn or relabeled since 1969 (Ref. [24]).

System affected by adverse drug reaction	Number of drugs affected (%)
Liver	74 (16)
Cardiovascular	40 (8.7)
Hematologic	39 (8.5)
Nervous system	36 (7.9)
Skin	34 (7.4)
Tumorigenicity	28 (6.1)
Urinary tract	28 (6.1)
Immunologic	27 (5.9)
Drug abuse	23 (5.0)
Psychiatric	17 (3.7)
Sensory systems	15 (3.3)
Gastrointestinal	12 (2.6)
Drug–drug interactions	11 (2.4)
Respiratory	11 (2.4)

Table 3
Likelihood of a cardiac clinical observation given the nonclinical observation.

Event	Sensitivity	Specificity	LR+	LR−
Arrhythmia	0.05	1.00	11.04	0.95
Electrocardiogram QT prolonged	0.15	0.99	10.70	0.86
Ventricular arrhythmias and cardiac arrest	0.06	0.99	7.22	0.95
Supraventricular arrhythmias	0.04	0.99	6.84	0.97
Vascular hypertensive disorders NEC	0.01	1.00	3.09	0.99

Table 4
Likelihood of a liver related clinical observation given the nonclinical observation.

Event	Sensitivity	Specificity	LR+	LR−
Hepatic function abnormal	0.01	1.00	25.63	0.99
Blood bilirubin increased	0.20	0.97	7.08	0.82
Aspartate aminotransferase increased	0.27	0.96	6.42	0.76
Alanine aminotransferase increased	0.47	0.93	6.29	0.57
Jaundice	0.01	1.00	5.13	0.99
Blood alkaline phosphatase increased	0.39	0.92	5.05	0.66
Hepatic necrosis	0.25	0.92	3.07	0.81
Hepatocellular damage and hepatitis NEC	0.26	0.91	2.98	0.81
Liver disorder	0.53	0.81	2.86	0.57
Hepatitis	0.04	0.98	2.16	0.98

Table 5
Likelihood of a kidney related clinical observation given the nonclinical observation.

Event	Sensitivity	Specificity	LR+	LR−
Blood creatinine increased	0.19	0.97	7.57	0.83
Renal failure and impairment	0.05	0.99	4.29	0.96
Nephropathies	0.35	0.88	2.98	0.73

Table 6
Likelihood of drug withdrawal given cardiac related clinical or nonclinical observations.

Event	Category	LR+	LR−
Arrhythmia	Preclinical	11.80	0.83
Electrocardiogram QT prolonged	Preclinical	8.94	0.80
Ventricular arrhythmias and cardiac arrest	Preclinical	8.44	0.87
C-reactive protein increased	Clinical	6.84	0.94
Supraventricular arrhythmias	Preclinical	5.76	0.95
Electrocardiogram QT prolonged	Clinical	5.32	0.55
Supraventricular arrhythmias	Clinical	3.22	0.52
Arrhythmia	Clinical	3.13	0.27
Ventricular arrhythmias and cardiac arrest	Clinical	2.58	0.53
Vascular hypertensive disorders NEC	Clinical	2.54	0.36

Table 7
Likelihood of drug withdrawal given liver related clinical or nonclinical observations.

Event	Category	LR+	LR−
Hepatic necrosis	Clinical	6.93	0.67
Hepatitis	Clinical	4.51	0.34
Jaundice	Clinical	4.17	0.35
Hepatic function abnormal	Clinical	3.52	0.76
Blood bilirubin increased	Clinical	3.52	0.54
Blood alkaline phosphatase increased	Preclinical	3.46	0.64
Hepatic necrosis	Preclinical	3.42	0.76
Blood alkaline phosphatase increased	Clinical	3.41	0.51
Alanine aminotransferase increased	Preclinical	3.33	0.57
Alanine aminotransferase increased	Clinical	3.03	0.44
Aspartate aminotransferase increased	Clinical	2.96	0.45
Hepatocellular damage and hepatitis NEC	Clinical	2.91	0.31
Blood bilirubin increased	Preclinical	2.81	0.90
Liver disorder	Clinical	2.71	0.88
Liver disorder	Preclinical	2.48	0.62
Hepatocellular damage and hepatitis NEC	Preclinical	2.00	0.84

which could lead to the conclusion that the preclinical test has no predictive value. The reason is that the overall prevalence of these events is low in clinical trials; only 56 drugs out of the 3815 are 'true positives' (have both nonclinical and clinical observations) for QT elongation, and only 12 for vascular hypertensive disorders.

3.2.2. Liver issues

For the liver related observations shown in Table 4, nonclinical observations in the class "Hepatic function abnormal" imply significant risk of those observations in humans with a likelihood ratio over 25.

Other nonclinical observations predictive for clinical observations include the markers alanine aminotransferase (ALAT) and aspartate aminotransferase (ASP), with positive likelihood ratio of 6.29 and 6.42. Unlike most liver observations, the lack of elevated ALAT results in a small decrease of risk of its observation in clinical studies. These statistics appears to vindicate the widespread use of ALAT and ASP as nonclinical biomarkers for liver issues in human. The same is true for bilirubin; if increased bilirubin is observed in animals for a given drug, the likelihood of it being observed in a clinical study is increased by 7.

Table 5 shows that for general nephropathy and other renal disorders, there is little statistical risk change due to a nonclinical observation, or lack of observation. The exception is increased creatinine, which increases the likelihood of a clinical observation by 7 times if observed in nonclinical study.

3.3. Predictive relationship of adverse events to drug withdrawal from market

Another metric of interest to the drug development process is the ability to identify specific observations as markers for either drug failure or drug withdrawal. Regulatory documents are not available for non-approved submissions, but 102 approved drugs have been withdrawn due to adverse events.

Table 6 shows statistics computed relating either clinical or nonclinical cardiac-related observations to drug withdrawal. The high likelihood of withdrawal if QT prolongation is observed is not surprising; many of the withdrawn drugs in this analysis were withdrawn specifically because of this effect. It is interesting that the nonclinical observation suggests more risk of withdrawal than the clinical, but this may also be the result of observational bias. The same is true of nonclinical arrhythmia observations; they suggest a mild increase in risk of withdrawal, where the clinical observation does not. The last interesting point is the low negative likelihood of withdrawal associated with lack of arrhythmia observations in clinical studies (0.27). Thus the lack of this observation statistically decreases the risk of drug failure.

Table 7 suggests that the lack of any hepatitis or jaundice observations in clinical studies lowers the risk of failure or withdrawal by a factor of 2.

3.4. Assessing the overall risk of drug withdrawal from market

Given the ability to compute likelihood of withdrawal based on adverse event observations, one can evaluate a drug's chance of withdrawal by taking the product of withdrawal risk for each

Table 8
Sample computation of total risk based on product of risk factors for trovafloxacin.

Event	Event category	LR+	LR−	Observed (+) or Not observed (−)	Resulting risk factor
Gastrointestinal ulceration and perforation	Preclinical	4.49	0.70	+	4.49
Gastrointestinal ulceration and perforation	Clinical	3.46	0.54	−	0.54
Anemia	Preclinical	4.53	0.86	−	0.86
Cerebrovascular accident	Clinical	3.69	0.62	−	0.62
Blood creatinine increased	Preclinical	3.05	0.88	+	3.05
Nephropathies	Preclinical	3.03	0.64	−	0.64
Nephropathies	Clinical	3.38	0.57	+	3.38
Renal failure and impairment	Preclinical	3.26	0.95	−	0.95
Injection and infusion site reactions	Preclinical	3.60	0.75	−	0.75
C-reactive protein increased	Clinical	6.84	0.94	−	0.94
Ventricular arrhythmias and cardiac arrest	Preclinical	8.44	0.87	+	8.44
Supraventricular arrhythmias	Preclinical	5.76	0.95	−	0.95
Supraventricular arrhythmias	Clinical	3.22	0.52	−	0.52
Arrhythmia	Preclinical	11.80	0.83	−	0.83
Arrhythmia	Clinical	3.13	0.27	+	3.13
Electrocardiogram QT prolonged	Preclinical	8.94	0.80	−	0.80
Electrocardiogram QT prolonged	Clinical	5.32	0.55	+	5.32
Hepatic function abnormal	Clinical	3.52	0.76	+	3.52
Hepatitis	Clinical	4.51	0.34	−	0.34
Jaundice	Clinical	4.17	0.35	−	0.35
Blood bilirubin increased	Clinical	3.52	0.54	−	0.54
Hepatic necrosis	Preclinical	3.42	0.76	+	3.42
Hepatic necrosis	Clinical	6.93	0.67	+	6.93
Blood alkaline phosphatase increased	Preclinical	3.46	0.64	−	0.64
Blood alkaline phosphatase increased	Clinical	3.41	0.51	−	0.51
Alanine aminotransferase increased	Preclinical	3.33	0.57	−	0.57
Alanine aminotransferase increased	Clinical	3.03	0.44	+	3.03
				Risk product	801.47

Table 9
Sample computation of total risk based on product of risk factors for hydrochlorothiazide.

Event	Event category	LR+	LR−	Test result	Resulting risk factor
Gastrointestinal ulceration and perforation	Preclinical	4.49	0.70	–	0.70
Gastrointestinal ulceration and perforation	Clinical	3.46	0.54	–	0.54
Anaemia	Preclinical	4.53	0.86	–	0.86
Cerebrovascular accident	Clinical	3.69	0.62	–	0.62
Blood creatinine increased	Preclinical	3.05	0.88	–	0.88
Nephropathies	Preclinical	3.03	0.64	–	0.64
Nephropathies	Clinical	3.38	0.57	–	0.57
Renal failure and impairment	Preclinical	3.26	0.95	–	0.95
Injection and infusion site reactions	Preclinical	3.60	0.75	–	0.75
C-reactive protein increased	Clinical	6.84	0.94	–	0.94
Ventricular arrhythmias and cardiac arrest	Preclinical	8.44	0.87	–	0.87
Supraventricular arrhythmias	Preclinical	5.76	0.95	–	0.95
Supraventricular arrhythmias	Clinical	3.22	0.52	–	0.52
Arrhythmia	Preclinical	11.80	0.83	–	0.83
Arrhythmia	Clinical	3.13	0.27	+	3.13
Electrocardiogram QT prolonged	Preclinical	8.94	0.80	–	0.80
Electrocardiogram QT prolonged	Clinical	5.32	0.55	–	0.55
Hepatic function abnormal	Clinical	3.52	0.76	–	0.76
Hepatitis	Clinical	4.51	0.34	–	0.34
Jaundice	Clinical	4.17	0.35	–	0.35
Blood bilirubin increased	Clinical	3.52	0.54	–	0.54
Hepatic necrosis	Preclinical	3.42	0.76	–	0.76
Hepatic necrosis	Clinical	6.93	0.67	–	0.67
Blood alkaline phosphatase increased	Preclinical	3.46	0.64	–	0.64
Blood alkaline phosphatase increased	Clinical	3.41	0.51	–	0.51
Alanine aminotransferase increased	Preclinical	3.33	0.57	–	0.57
Alanine aminotransferase increased	Clinical	3.03	0.44	–	0.44
				Risk product	0.00004

adverse event category. The two drugs trovafloxacin and hydrochlorothiazide are compared. While hydrochlorothiazide is well tolerated, the use of trovafloxacin has been restricted due to liver toxicity [26].

Table 8 shows an example computation of total risk for the drug trovafloxacin based on the Bayesian statistics and the risk factors associated with the observation, or lack of observation of statistically significant events (those with positive likelihood ratios, LR+, greater than 3.0). If the observation is made, the resulting risk factor is drawn from the LR+ column, the positive likelihood. If the observation is not made the risk factor is drawn from the LR− column, the negative likelihood. The total risk is the product of these factors. This analysis computes the likelihood of failure of trovafloxacin to be over 800. That is, this drug is 800 times more likely to fail than the "typical" approved drug. In contrast, Table 9 shows the risk factor for hydochlorothiazide, a drug with comparatively few serious reported adverse events. The computed risk factor for hydrochlorothiazide is 0.00004. To extend this work one could compute this value over all events for all drugs to create a baseline for evaluation of ongoing clinical programs.

The advantage of the likelihood ratio over the positive predictive value is that it can be used to evaluate drug classes to give an accurate probability of withdrawal. For example, the prevalence of withdrawal among the fluoroquinolone antibiotics is 0.25 (4 withdrawn of 16 approved), or odds of 1 in 3 ($0.25/(1-0.25)$). The odds of withdrawal of a fluoroquinolone antibiotic, if QT elongation is seen in a preclinical study, is the pretest odds times the likelihood ratio, $8.94 * 0.333$, or 2.98. This is close to 3–1 odds of being withdrawn, which is equivalent to a probability of 0.75 ($3/(3+1)$) – far greater than the initial probability of 0.25 – if QT elongation is observed in a nonclinical study of a drug in this class.

4. Conclusions

In this study we proposed a framework based on likelihood ratio computed using 2×2 contingency matrices for studying the relationship between nonclinical and clinical adverse events in order to create robust analyses to understand the implications of nonclinical adverse events to clinically observed adverse events. The framework was also used to understand the implication of various nonclinical and clinical adverse events for drug withdrawal.

From this study we learned that some adverse events observed in nonclinical studies are strongly associated with the observation in corresponding clinical studies. For example nonclinical observation of cardiac arrhythmia and QT interval prolongation is strongly suggestive that it will be observed in clinical studies. Others, such as hepatitis, are far less strongly associated and many more. such as cholestasis, do not have statistically significant relationships between nonclinical and clinical observations. While observation of many nonclinical adverse events appear to be related to a high risk of seeing the same event in humans, little statistical inference for clinical results could be made for most cases when a particular adverse event was not observed.

In addition analysis of the relationship between both nonclinical and clinical adverse events and drug withdrawal showed that some adverse events are much more strongly associated with drug withdrawal than others, particularly arrhythmias and hepatic necrosis. The idea of using the product of the individual risk factors as an overall metric was tested on hydrochlorothiazide and trovafloxacin and for those two cases the product was consistent with the known safety of the first, and withdrawal of the latter drug. The concept of using the product of the individual risk factors for withdrawal as an overall metric was tested on hydrochlorothiazide and trovafloxacin and for those two cases the product was consistent with the known safety of the first, and withdrawal of the latter drug.

Acknowledgments

The author acknowledges Sherri Matis-Mitchell (Elsevier Life Sciences) and Pieder Caduff (Elsevier REPSA) for discussions and comments.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jbi.2015.02.008.

References

[1] Guidance for industry nonclinical safety evaluation of drug or biologic combinations. United States Food Drug Adm; 2006, <http://www.fda.gov/OHRMS/DOCKETS/98fr/05d-0004-gdl0002.pdf>.

[2] Jamieson RC. Eyelash-dye (Lash-Lure) dermatitis with conjunctivitis. J Am Med Assoc 1933;20:1560.

[3] Wax PM. Federal food, drug and cosmetic act. Ann Intern Med 1938;122(1995):456–61. http://dx.doi.org/10.1097/00043764-199607000-00005.

[4] Taussig HB. A study of the German outbreak of phocomelia. J Am Med Assoc 1962;180:1106–14. http://dx.doi.org/10.1001/jama.1962.03050260028005.

[5] Gad S. Animal models in toxicology. Boca Raton: Taylor & Francis Group; 2007.

[6] Shanks N, Greek R, Greek J. Are animal models predictive for humans? Philos Ethics Humanit Med 2009;4:2. http://dx.doi.org/10.1186/1747-5341-4-2.

[7] Goldsmith MA, Slavik M, Carter SK. Quantitative prediction of drug toxicity in humans from toxicology in small and large animals. Cancer Res 1975;35:1354–64.

[8] Fletcher AP. Drug safety tests and subsequent clinical experience. J R Soc Med 1978;71:693–6.

[9] Olson H, Betton G, Robinson D, Thomas K, Monro A, Kolaja G, et al. Concordance of the toxicity of pharmaceuticals in humans and in animals. Regul Toxicol Pharmacol 2000;32:56–67. http://dx.doi.org/10.1006/rtph.2000.1399.

[10] Perel P, Roberts I, Sena E, Wheble P, Briscoe C, Sandercock P, et al. Comparison of treatment effects between animal experiments and clinical trials: systematic review. BMJ 2007;334:197. http://dx.doi.org/10.1136/bmj.39048.407928.BE.

[11] Bolton C. The translation of drug efficacy from in vivo models to human disease with special reference to experimental autoimmune encephalomyelitis and multiple sclerosis. Inflammopharmacology 2007;15:183–7. http://dx.doi.org/10.1007/s10787-007-1607-z.

[12] Zbinden G. Predictive value of animal studies in toxicology. Regul Toxicol Pharmacol 1991;14:167–77. http://dx.doi.org/10.1016/0273-2300(91)90004-F.

[13] Heywood R. Clinical toxicity–could it have been predicted? post-marketing experience. In: Lumley CE, Walker S, editors. Animal toxicity studies: their relevance for man. Quay, Lancaster; 1990. p. 57–67.

[14] Tabacova S, Kimmel CA, Wall K, Hansen D, Jackman GP, Iakovidis D, et al. Atenolol developmental toxicity: animal-to-human comparisons. Birth Defects Res Part A Clin Mol Teratol 2003;67:181–92.

[15] Wall RJ, Shani M. Are animal models as good as we think? Theriogenology 2008;69:2–9. http://dx.doi.org/10.1016/j.theriogenology.2007.09.030.

[16] Combes R, Grindon C, Cronin MTD, Roberts DW, Garrod JF. Integrated decision-tree testing strategies for mutagenicity and carcinogenicity with respect to the requirements of the EU REACH legislation. ATLA Altern Lab Anim 2008;36:43–63.

[17] Pharmapendium. Reed Elsevier Properties SA; 2014. <https://pharmapendium.com>.

[18] Brown EG, Wood L, Wood S. The medical dictionary for regulatory activities (MedDRA). Drug Saf 1999;20:109–17. http://dx.doi.org/10.2165/00002018-199920020-00002.

[19] Mosby's Drug Consult 13th ed. Mosby, St. Louis, MO; 2003.

[20] Dukes MNG. Meyler's side effects of drugs. 12th ed. Amsterdam, Amsterdam: Elsevier Science Publishers BV; 1992.

[21] Grimes DA, Schulz KF. Refining clinical diagnosis with likelihood ratios. Lancet 2005;365:1500–5. http://dx.doi.org/10.1016/S0140-6736(05)66422-7.

[22] Chien PFW, Khan KS. Evaluation of a clinical test. II: Assessment of validity. Br J Obstet Gynaecol 2001;108:568–72. http://dx.doi.org/10.1016/S0306-5456(00)00128-5.

[23] Core R. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2012.

[24] Aronson JK. Drug withdrawals because of adverse effects. Side Eff Drugs Annu 2008;30. http://dx.doi.org/10.1016/S0378-6080(08)00064-0.

[25] Morita H, Wu J, Zipes DP. The QT syndromes: long and short. Lancet 2008;372:750–63. http://dx.doi.org/10.1016/S0140-6736(08)61307-0.

[26] Peters TS. Do preclinical testing strategies help predict human hepatotoxic potentials? Toxicol Pathol 2005;33:146–54. http://dx.doi.org/10.1080/01926230590522121.